yang_seonhyeHW11

```r
library(alr4, quietly = T, logical.return = FALSE, warn.conflicts = FALSE)
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```r
library(data.table)
attach(fuel2001, warn.conflicts = F)
head(fuel2001)
```

```
##     Drivers    FuelC Income  Miles      MPC      Pop  Tax
## AL  3559897  2382507  23471  94440 12737.00  3451586 18.0
## AK   472211   235400  30064  13628  7639.16   457728  8.0
## AZ  3550367  2428430  25578  55245  9411.55  3907526 18.0
## AR  1961883  1358174  22257  98132 11268.40  2072622 21.7
## CA 21623793 14691753  32275 168771  8923.89 25599275 18.0
## CO  3287922  2048664  32949  85854  9722.73  3322455 22.0
```

```r
model<- lm(FuelC~Tax+Drivers+Income)
```

Question 1

```r
rmodel<-rstudent(model)
rmodel
```

```
##           1           2           3           4           5           6
## -0.46986746 -0.82491434 -0.22552020 -0.26084510  0.70545119 -0.05630035
##           7           8           9          10          11          12
##  0.19227116  0.24295695  0.30391038 -3.26347861  1.14716565 -0.56496830
##          13          14          15          16          17          18
## -0.06721422 -0.37602240  0.41402139  0.15750910 -0.10053959  0.04140851
##          19          20          21          22          23          24
##  0.36507551 -0.23761917  0.75825045 -0.29709900  0.46684060  1.42180335
##          25          26          27          28          29          30
## -0.07658889  0.55145654 -0.06313120  0.02320195  0.26378418  0.25899886
##          31          32          33          34          35          36
##  0.16981675 -0.38521666 -4.70659975  0.32209073 -0.16259656 -0.31755855
##          37          38          39          40          41          42
##  0.15315873 -0.35581807 -0.82241372  0.34623857  0.17893242 -0.03417366
##          43          44          45          46          47          48
## -0.12936141  5.74349084 -0.25316822 -0.19927711  1.02258646 -0.22841001
##          49          50          51
## -0.33722937  0.18520584 -0.28225433
```

Question 2

```r
# Find critical value
cVal <- abs(qt(0.05/2, nobs(model) - 1))
rmodel[abs(rmodel) > cVal]
```

```
##        10        33        44
```

```
## -3.263479 -4.706600  5.743491
```

```r
# Find Bonferroni-adjusted critical value
cValBonfer <- abs(qt(0.05/(2*nobs(model)), df=df.residual(model)-1, lower.tail=FALSE))
rmodel[abs(rmodel) > cValBonfer]
```

```
##        33        44
## -4.706600  5.743491
```

The mean shift test shows that number 10, 33 and 44 are outliers and the Bonferroni test shows that number 33 and 44 are outliers.

Question 3

```r
cdist <- cooks.distance(model)
cdist[cdist >= 1]
```

```
## named numeric(0)
```

Looking at Cook's distance, we can see that there is no influential points.

Question 4

We already know that $H = X(X^TX)^{-1}X^T$ if we multiply both sides by X we get the following:

$$H = X(X^TX)^{-1}X^T$$
$$HX = (X(X^TX)^{-1}X^T)X$$

$$HX = X(X^TX)^{-1}(X^TX)$$
$$X^TX^{-1}(X^TX) = I$$
$$HX = XI$$

Therefore, $HX = XI$ is equal to $HX = X$

Now, in order to show that $X^TH = X^T$ we can multiply $H = X(X^TX)^{-1}X^T$ by X^T on both sides:

$$X^TH = X^TX(X^TX)^{-1}X^T$$
$$X^TX(X^TX)^{-1} = I$$

$$X^TH = X^TX(X^TX)^{-1}X^T = X^TI$$

$$X^TI = X^T$$

$$X^TH = X^T$$

Question 5

Since the model has an intercept, the column of the matrix will look like this:

$$X = [x_0, x_1, x_2 ... x_n]$$

where the rows of $x_0 = [1, 1, 1..]$ because the model has an intercept.

From question 4,

$$X^T H = X^T$$

$$X^T H = X^T = [x_0^T, x_1^T, x_2^T ... x_n^T]' H = [x_0^T, x_1^T, x_2^T ... x_n^T]'$$

Therefore, $x_0^T H = x_0^T$ where

$$(1, 1, 1..)' \begin{bmatrix} h_1 & h_2 & h_3 ... h_n & \\ . & . & . & . \end{bmatrix} = (1, 1, 1..)'$$

$$\sum_{i=1}^{n} h_{i1} \sum_{j=1}^{n} h_{j1} ... \sum_{i=1}^{n} h_{in} \sum_{j=1}^{n} h_{jn} = (1, 1, 1....1)$$

$$\sum_{i=1}^{n} h_{ij} = 1$$

This means that for all the elements in $j = 1n$

Now from $HX = X$ we can write this as

$$\begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & h_{13} & ... & h_{1n} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ h_{n1} & . & . & . & h_{nn} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \sum_{j=1}^{n} h_{1j} \\ \sum_{j=1}^{n} h_{2j}. \\ . \\ . \\ \sum_{j=1}^{n} h_{nj} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}$$

$$\sum_{j=1}^{n} h_{ij} = 1$$

This means that for all the elements in $i = 1n$

Therefore,

$$\sum_{i=1}^{n} h_{ij} = \sum_{j=1}^{n} h_{ij} = 1$$